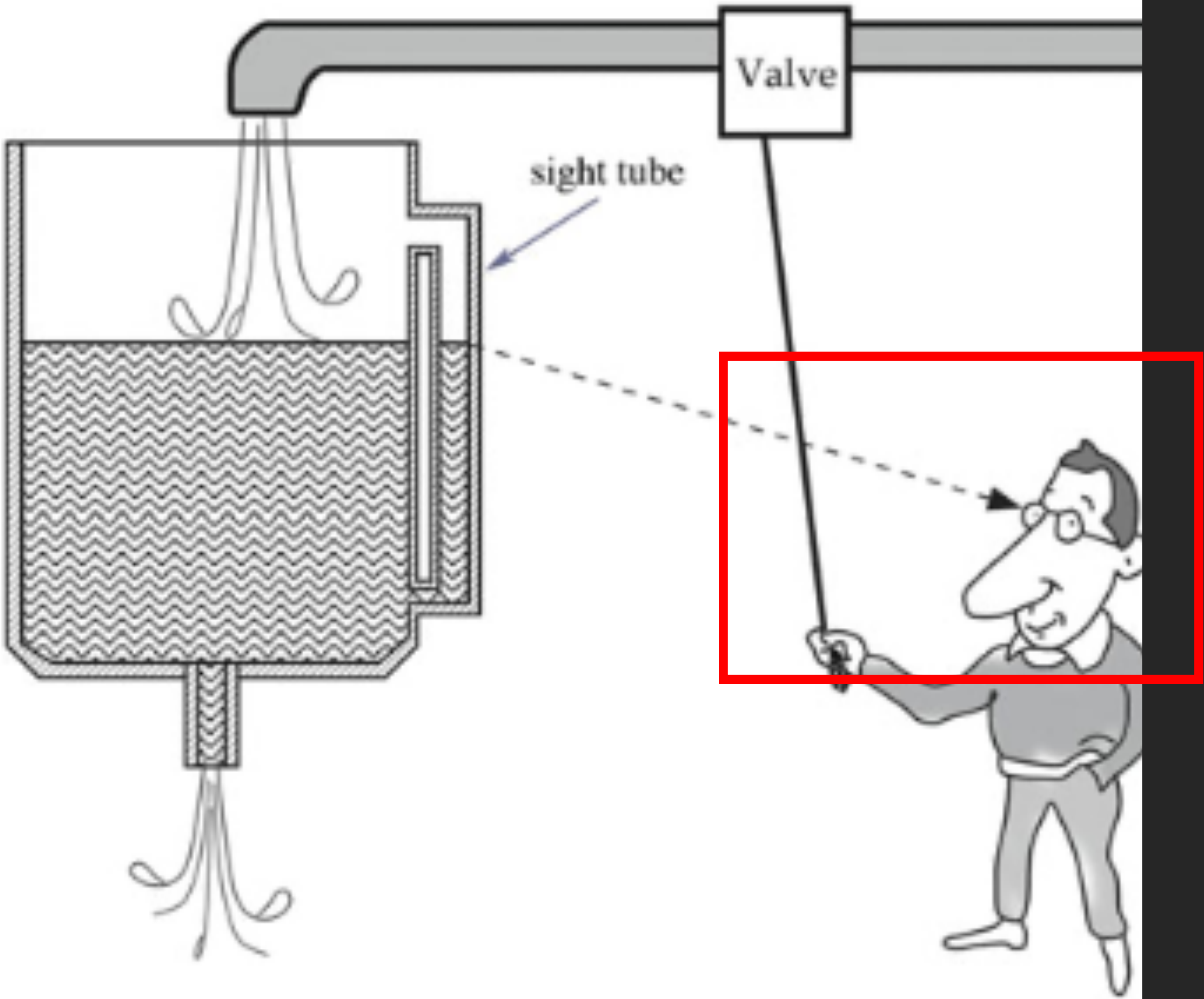
Tópicos Especiais em Hidráulica e Saneamento: Sensores e Novas Tecnologias para a Melhoria da Qualidade de Água com Monitoramento em Tempo Real

SHS5964

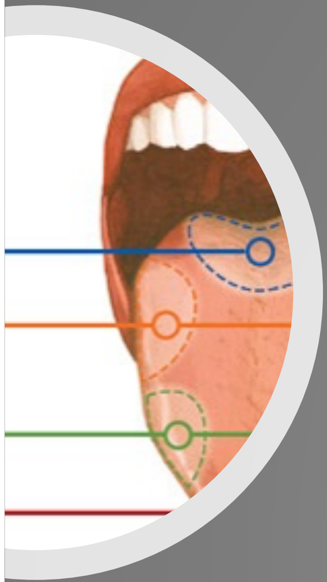
Prof. Dr. Filippo Ghiglieno

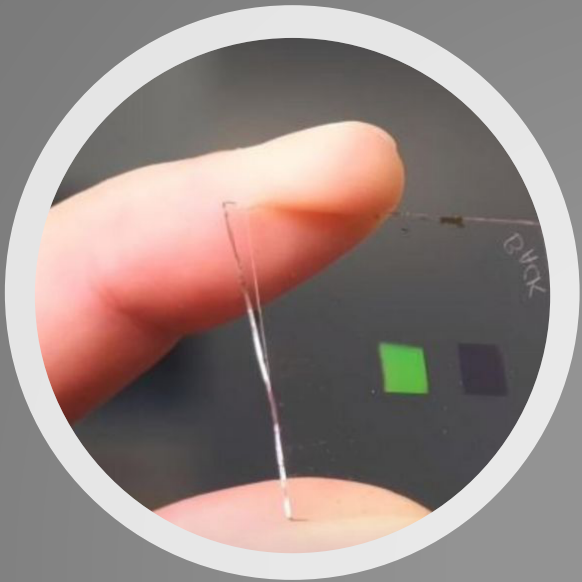
filippo.ghiglieno@df.ufscar.br

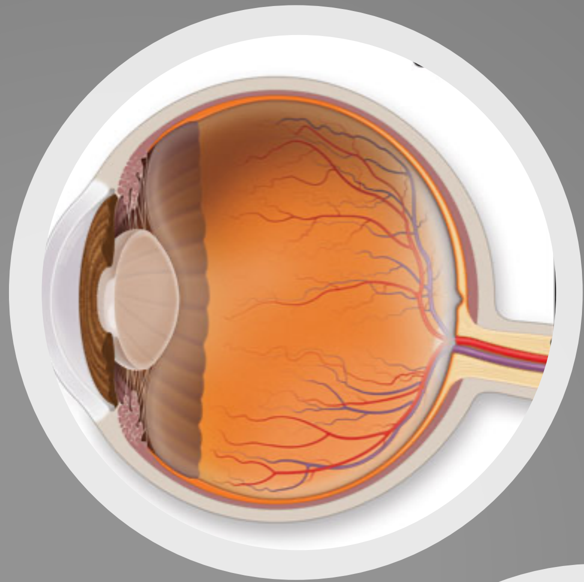
(DF/UFSCar)



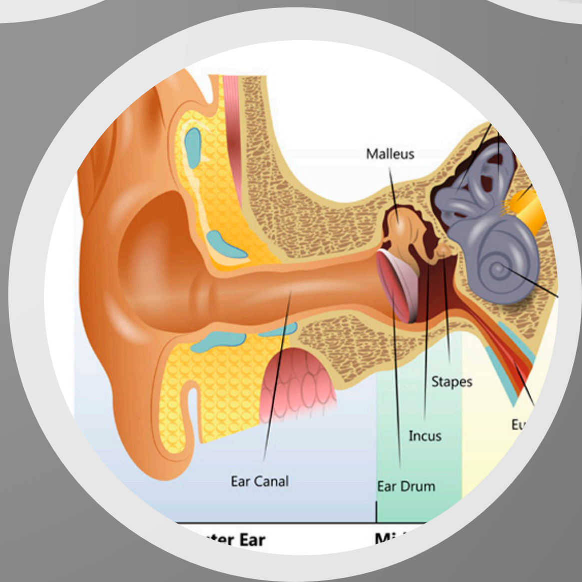
A sensor is
«a device that receives and responds to a signal stimulus»

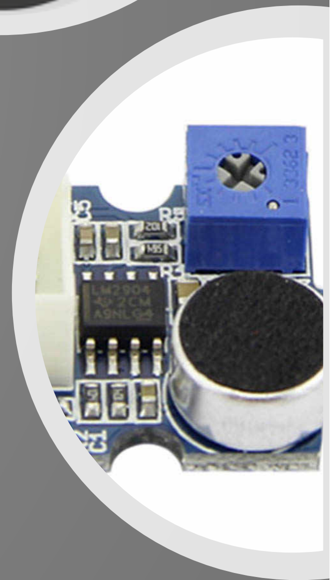






- Natural sensors ⇒ electrochemical signal
- Man made sensors ⇒ electrical signal





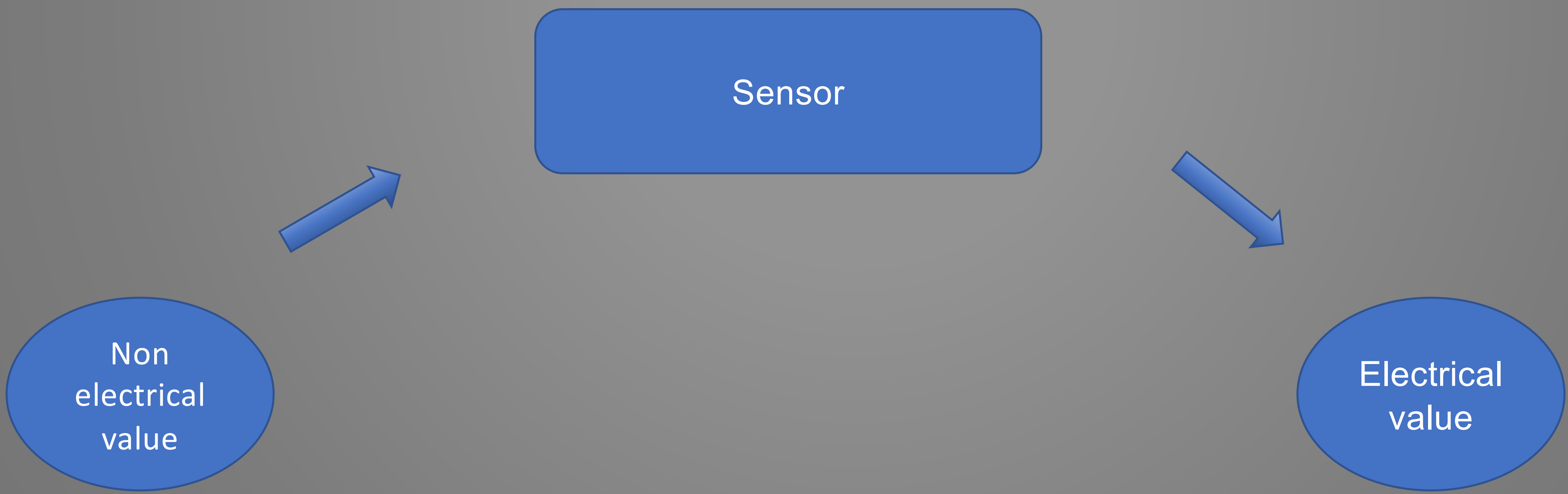
A sensor is
«a device that receives a stimulus and responds with an electrical signal»

The stimulus

is the quantity, the property or the condition that is received and converted into electrical signal

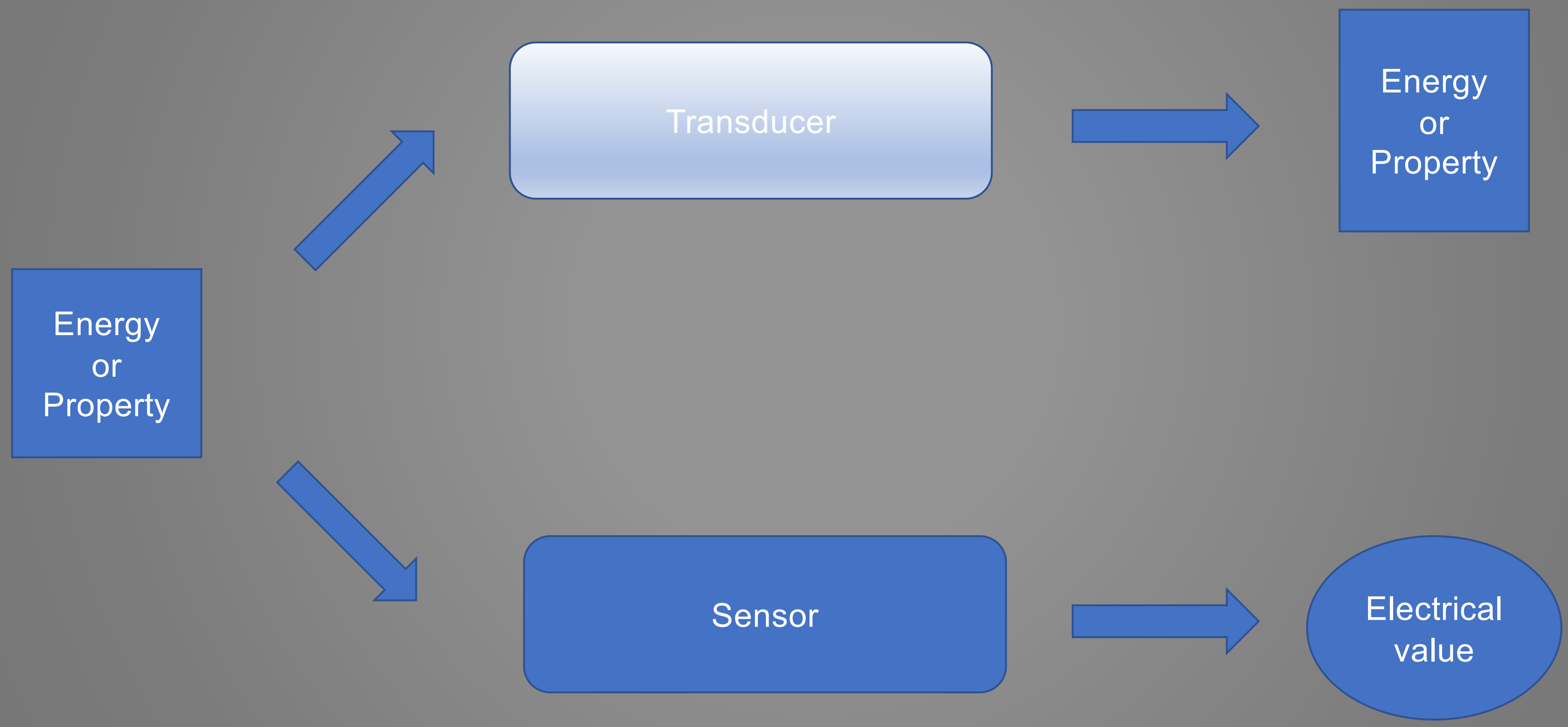
The measurand

TRANSLATOR

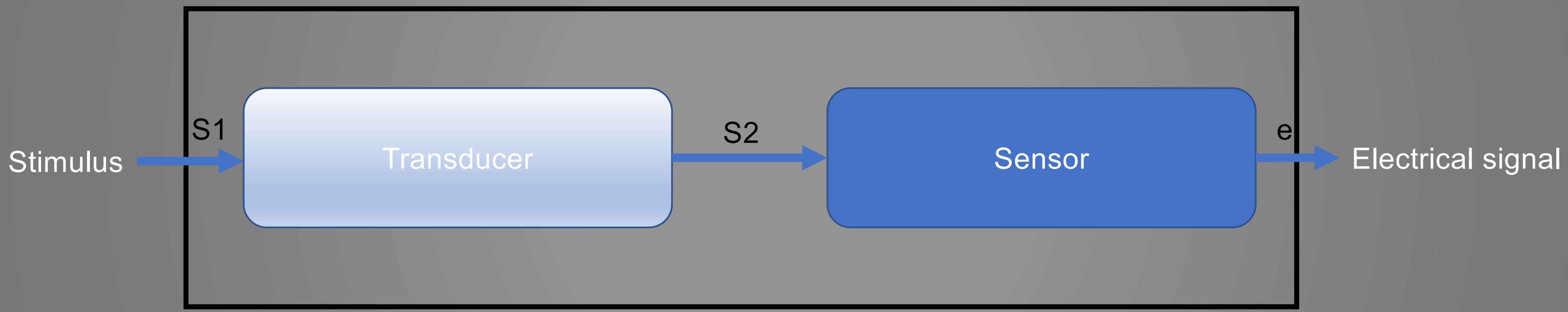


Any sensor is an energy converter:

the energy transfers between the object of measurement to the sensor

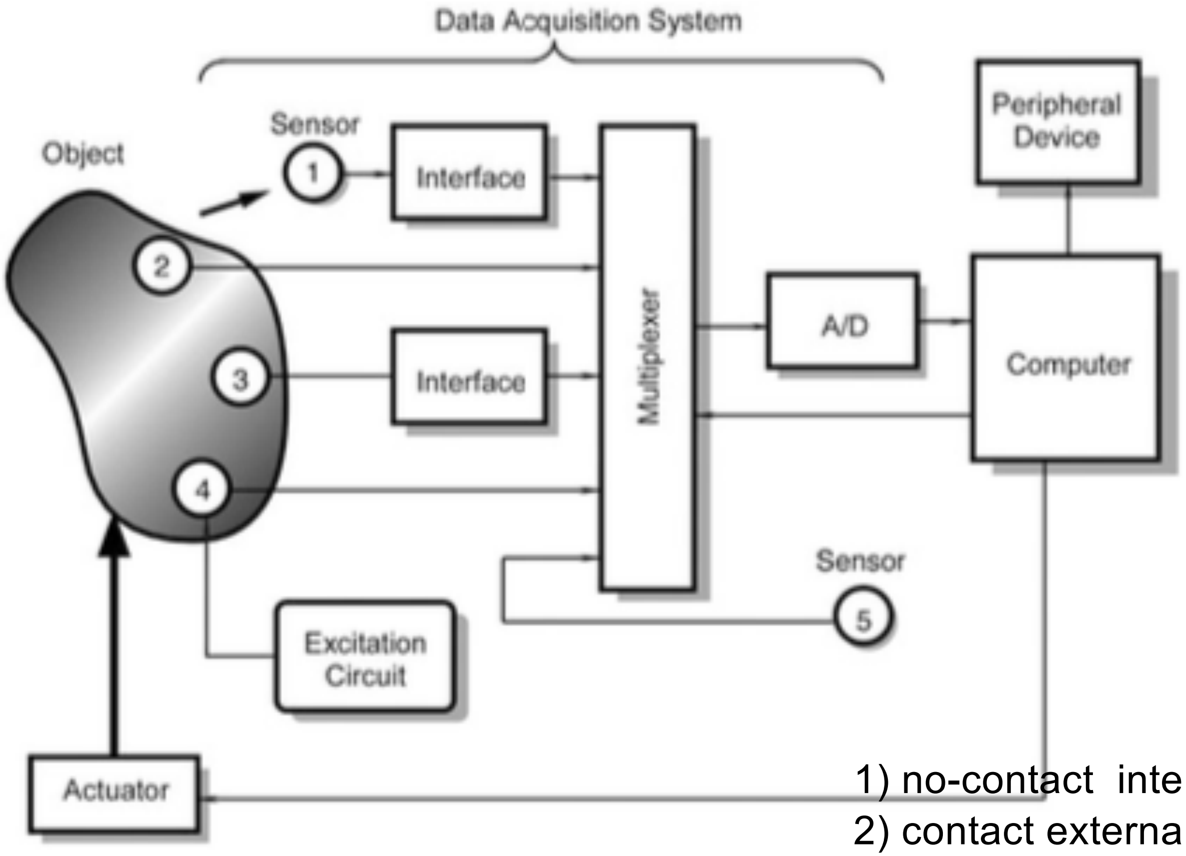


Hybrid or complex sensor



A sensor does not function by itself; it is always part of a larger system that may incorporate many other detectors, signal conditioners, processors, memory devices, data recorders, and actuators.

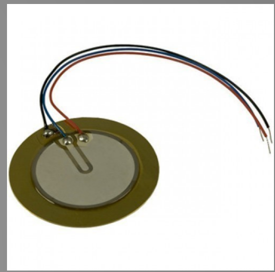
Data acquisition and control device : the position of sensors



1) no-contact internal passive (with signal conditioner) sensor;
2) contact external passive sensor;
3) contact external passive (with signal conditioner) sensor;
4) contact external active sensor;
5) internal passive sensor.

All sensors may be of two kinds: **passive and active**.

A passive sensor does not need any additional energy source. It generates an electric signal in response to an external stimulus. That is, the input stimulus energy is converted by the sensor into the output signal: thermocouple, a photodiode, and a piezoelectric sensor.



The active (parametric) sensors require external power for their operation, which is called an excitation signal. That signal is modified (modulated) by the sensor to produce the output signal. It can be stated that a sensor's parameter modulates the excitation signal and that modulation carries information of the measured value. For example, a thermistor is a temperature-sensitive resistor. It does not generate any electric signal, but by passing electric current (excitation signal) through it its resistance can be measured by detecting variations in current and/or voltage across the thermistor. These variations (presented in ohms) directly relate to temperature through a known transfer function.

Sensor specification
Sensitivity
Accuracy
Speed of response
Overload characteristics
Hysteresis
Operating life
Cost, size, weight
Stimulus range (span)
Resolution
Selectivity
Environment conditions
Linearity
Dead band
Output Format

Conversion phenomena	
Physical	Thermoelectric
	Photoelectric
	Photomagnetic
	Magnetoelectric
	Electromagnetic
	Thermoelastic
	Thermo-optic
	Photoelastic
Chemical	Chemical transformation
	Physical transformation
	Electrochemical process
	Spectroscopy
Biological	Biochemical transformation
	Physical transformation
	Effect on test organism
	Spectroscopy

Stimulus	
Acoustic	Wave amplitude
	Spectrum polarization
	Wave velocity
Biological	Biomass
Chemical	Identities
	Concentration
	State
Electric	Charge, Current
	Potential, Voltage
	Electric field (amplitude, phase, polarization, spectrum)
	Conductivity
	Permittivity
Magnetic	Magnetic field (amplitude, phase, polarization, spectrum)
	Magnetic flux
	Permeability

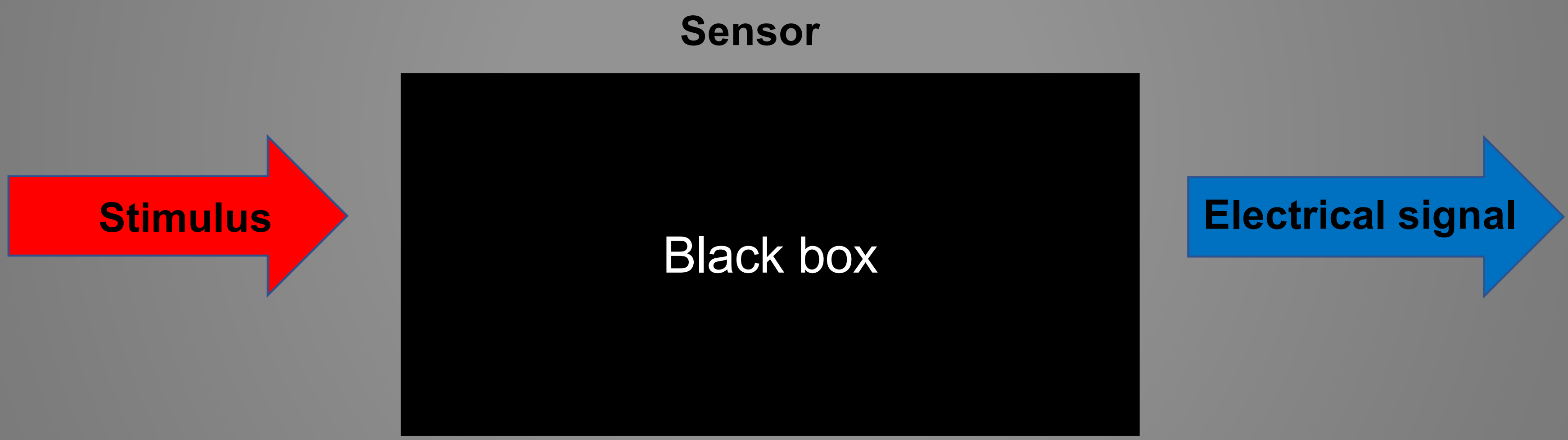
Stimulus	
Optical	Wave amplitude, phase, polarization, spectrum
	Wave velocity
	Refractive index
	Emissivity, refractivity, absorption
Mechanical	Position
	Acceleration
	Force
	Stress, pressure
	Strain
	Mass, density
	Moment, torque
	Speed of flow, rate of mass transport
	Shape, roughness, orientation
	Stiffness, compliance
	Viscosity
	Crystallinity, structural integrity

Stimulus	
Radiation	Type
	Energy
	Intensity
Thermal	Temperature
	Flux
	Specific heat
	Thermal conductivity

Unit of measurements

Quantity	Name	Symbol	Defined by
Length	mater	m	… the length of the path travelled by light in vacuum in 1/2999792458 of a second (1983)
Mass	kilogram	kg	…after a platinum-iridium prototype (1889)
Time	second	s	…the duration of 9192631770 periods of the rotation corresponding to the transition between the 2 hyperfine levels of the ground state of the cesium 133 atom (1967)
Electric current	ampere	A	Force equal to $2x10^{-7}$ N/m of length exerted on two parallel conductors in vacuum when they carry the current (1946)
Thermodynamic temperature	kelvin	K	The fraction 1/273.16 of the thermodynamic temperature of the triple point of water (1967)
Amount of substance	mole	mol	. . .the amount of substance which contains as many elementary entities as there are atoms in 0.012 kg of carbon 12 (1971)
Luminous intensity	candela	cd	. . .the amount of substance which contains as many elementary entities as there are atoms in 0.012 kg of carbon 12 (1971)
Plane angle	radian	rad	(supplemental unit)
Solid angle	steradian	sr	(supplemental unit)

Since most of *stimuli* are not electrical, from its input to the output a sensor may perform several signal conversion steps before it produces and outputs an electrical signal.

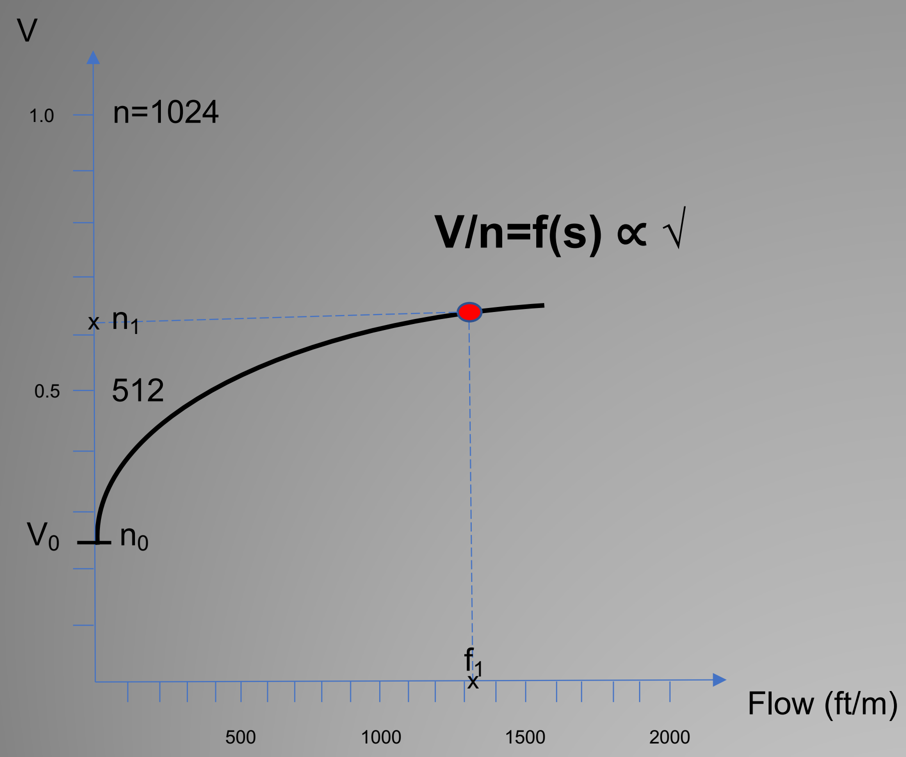


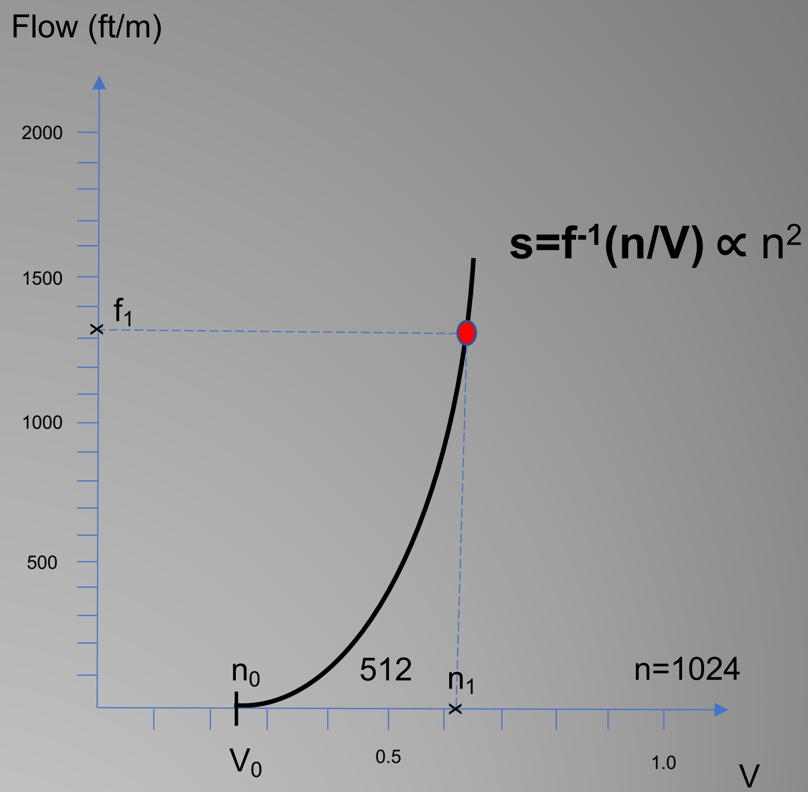
Static transfer function

$s \longrightarrow f(s) \longrightarrow E$

Thermo-anemometer: transfer function

n/V=f(s) ➔ transfer function [sensor + electronic circuit (e.g. ADC converter)]





Functional approximation

curve fitting of the experimentally observed values into the approximating function

Linear transfer function

$$E = E_0 + B(s - s_0)$$

$$s = s_0 + \frac{E - E_0}{B}$$

Exponential transfer function

$$E = Ae^{k \cdot s}$$

$$s = \frac{1}{k}\ln(E)$$

Logarithmic transfer function

$$E = A + B \cdot \ln(s)$$

$$s = e^{\frac{E-A}{B}}$$

Power transfer function

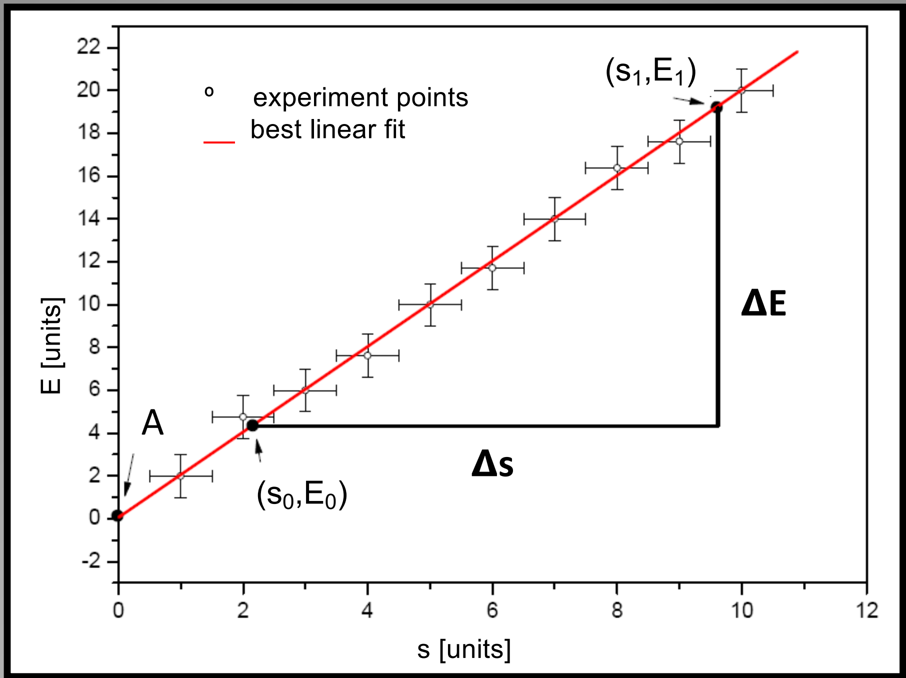
$$E = A + Bs^k$$

$$s = \sqrt[k]{\frac{E - A}{B}}$$

Linear transfer function
⇳
Linear regression – least squares

$$E = E_0 + B(s - s_0)$$

$$s = s_0 + \frac{E - E_0}{B}$$



$$E = A + Bs$$

$$s = -\frac{A}{B} + \frac{E}{B}$$

$$A = \frac{\sum_i E_i \sum_j s_j^2 - \sum s_i \sum_j E_j s_j}{k \sum_i s_i^2 - \left(\sum_i s_i\right)^2}$$

$$B = \frac{k \sum_i E_i s_i - \sum_i E_i \sum_j s_j}{k \sum_i s_i^2 - \left(\sum_i s_i\right)^2}$$

Non-linear transfer function
⇕
Power series

$$E = Ae^{ks} \approx \left(1 + ks + \frac{k^2}{2!}s^2 + \frac{k^3}{3!}s^3 + O(s^4)\right)$$

$E \approx a_2 s^2 + a_1 s + a_0$

s$\approx A_2 E^2 + A_1 E + A_0$

⇔

$E \approx b_3 s^3 + b_2 s^2 + b_1 s + b_0$

s$\approx B_3 E^3 + B_2 E^2 + B_1 E + B_0$

Second order polynomial often may yield a fit of sufficient accuracy when applied to *relatively narrow range of input stimuli* and *the transfer function is monotonic* (no ups and downs)

Sensitivity

The coefficient B in

$$E = A + Bs$$

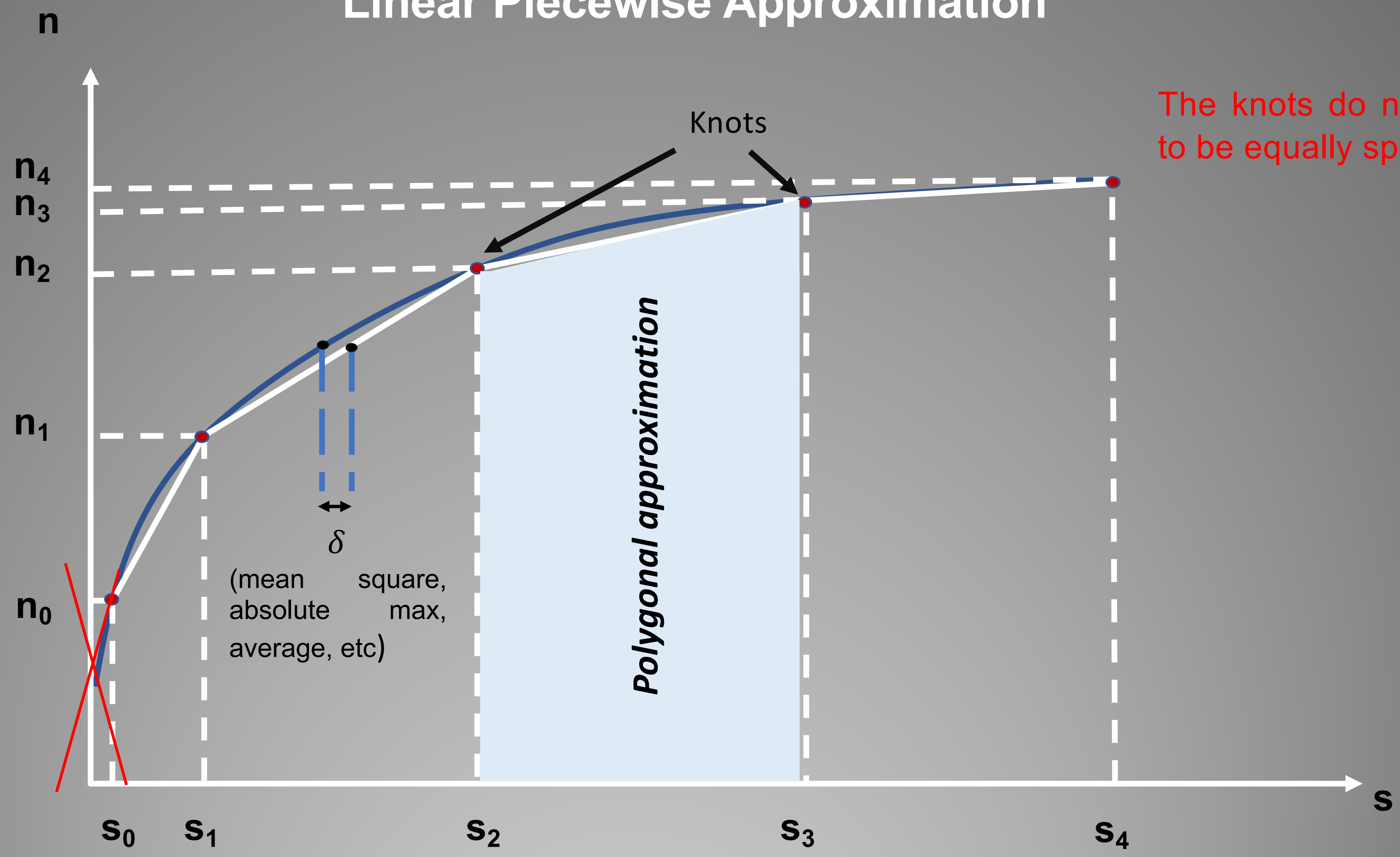
$$E = E_0 + B(s - s_0)$$

is called **sensitivity**. For a nonlinear transfer function, sensitivity is not a fixed number, as would be the case in a linear transfer function. A nonlinear transfer function exhibits different sensitivities at different points in intervals of stimuli. In the case of nonlinear transfer functions, sensitivity is defined as a first derivative of the transfer function at the particular stimulus s_i:

$$b_i(s_i) = \frac{dE_i(s_i)}{ds_i} = \frac{\Delta E_i}{\Delta s_i}$$

where, Δs_i is a small increment of the input stimulus and ΔE_i is the corresponding change in the sensor output E.

Linear Piecewise Approximation



The knots do not need to be equally spaced

The larger the number of knots the smaller the error. They should be closer to each other where linearity is high, and farther apart where nonlinearity is small.

Spline interpolation

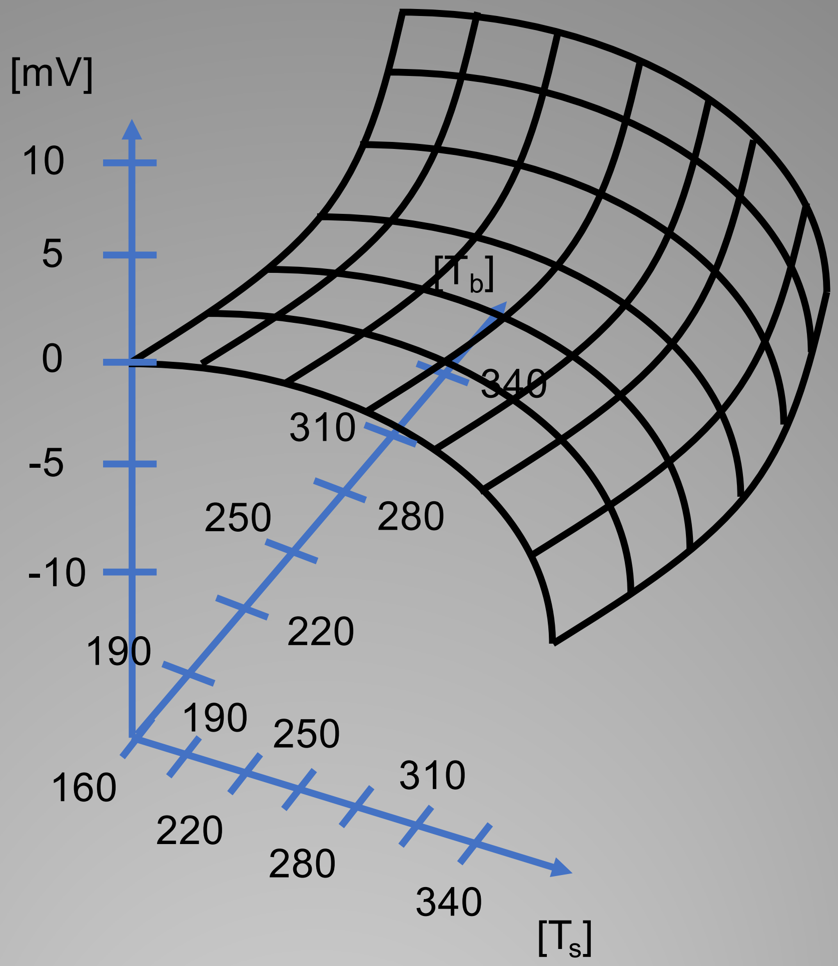
Approximations by higher order polynomials (third order and higher) have some disadvantages; the selected points at one side of the curve make strong influence on the remote parts of the curve. This deficiency is resolved by the **spline method** of approximation. In a similar way to a linear piecewise interpolation, the spline method is using different third-order polynomial interpolations between the selected experimental points called knots. It is a curve between two neighbouring knots and then all curves are "glued" together to obtain a smooth combined curve fitting. Not necessarily it should be a third-order curve, it can be as simple as the first-order (linear) interpolation. A linear spline interpolation (first order) is the simplest form and is equivalent to a linear piecewise approximation as described in the previous slide.

The spline interpolation can utilize polynomials of different degrees, yet the most popular being cubic (third order) polynomials. Curvature of a line at each point is defined by the second derivative. This derivative should be computed at each knot. If the second derivatives are zero, the cubic spline is called "**relaxed**" and it is the choice for many practical approximations. Spline interpolation is the efficient technique when it comes to an interpolation that preserves smoothness of the transfer function. However, simplicity of the implementation and the computational costs of a spline interpolation should be taken into account particularly in a tightly controlled microprocessor environment.

Multidimensional transfer function

A sensor transfer function may depend on more than one input variable. That is, the sensor's output may be a function of several stimuli. One example is a humidity sensor whose output depends on two input variables, relative humidity and temperature. Another example is the transfer function of a thermal radiation (infrared) sensor. This function has two arguments:

- the absolute temperature of an object of measurements [T_b];
- The absolute temperature of the sensing element [T_s].



$$V = G(T_b^4 - T_s^4)$$

$$G = 10^{-12}$$

Calibration

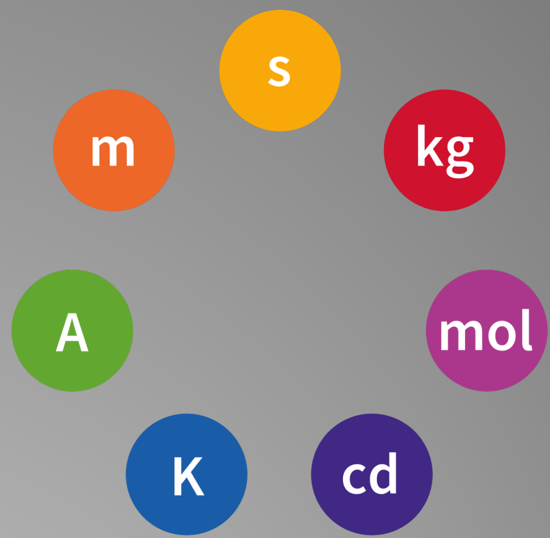
If tolerances of a sensor and interface circuit (signal conditioning) are broader than the required overall accuracy, a calibration of the sensor or, preferably, a combination of a sensor and its interface circuit is required for minimizing errors. In other words, a calibration is required whenever a higher accuracy is required from a less accurate sensor. For example, if one needs to measure temperature with accuracy, say 0.1^{o}C, while the available sensor is rated as having accuracy of 1^{o}C, it does not mean that the sensor cannot be used. Rather this particular sensor needs calibration. That is, its unique transfer function should be determined. This process is called calibration.

Calibration

A calibration requires application of several precisely known stimuli and reading the corresponding sensor responses. These are called the calibration points whose input–output values are the point coordinates. In some lucky instances only one pair is required, while **typically 2–5 calibration points are needed** to characterize a transfer function with a higher accuracy. After the unique transfer function is established, any point in between the calibration points can be determined.

Calibration

To produce the calibration points, a standard reference source of the input stimuli is required. The reference source should be well maintained and periodically checked against other established references, preferably traceable to a national standard, for example a reference maintained by NIST in the U.S.A. It should be clearly understood that the calibration accuracy is directly linked to accuracy of a reference sensor that is part of the calibration equipment. A value of uncertainty of the reference sensor should be included in the statement of the overall uncertainty.



Before calibration, either a mathematical model of the transfer function has to be known or a good approximation of the sensor's response over the entire span shall be found. In a great majority of cases, such functions are smooth and monotonic. Very rarely they contain singularities and if they do, such singularities are the useful phenomena that are employed for sensing.

Calibration

Before calibration, either a mathematical model of the transfer function has to be known or a good approximation of the sensor's response over the entire span shall be found. In a great majority of cases, such functions are smooth and monotonic. Very rarely they contain singularities and if they do, such singularities are the useful phenomena that are employed for sensing (an ionizing particle detector is an example).

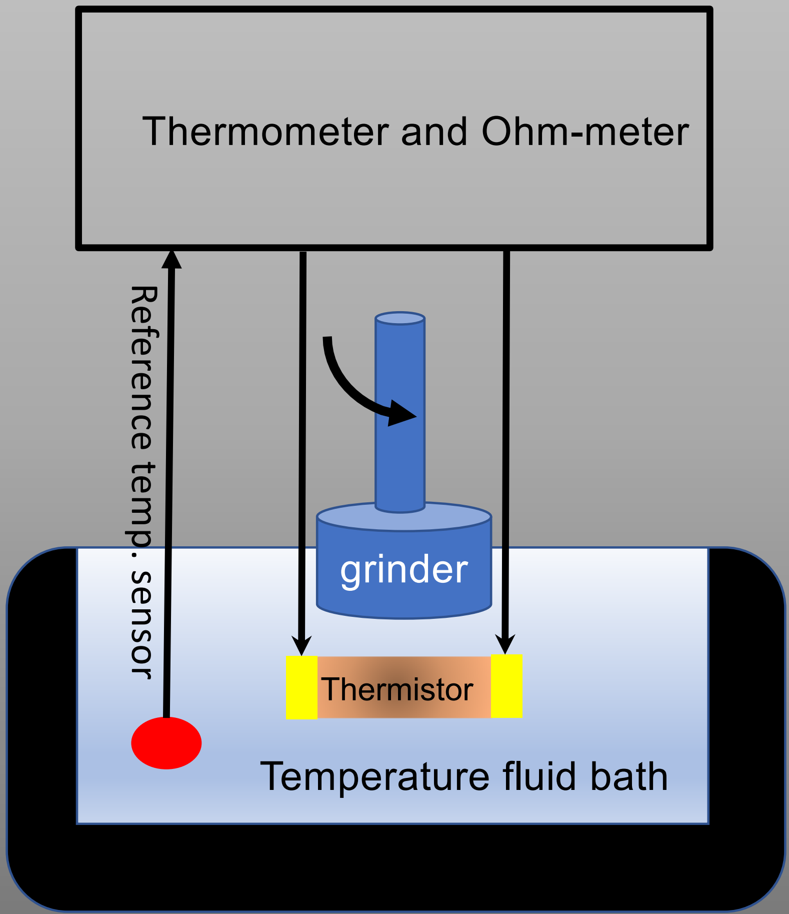
Calibration of a sensor can be done in several possible ways, some of which are the following:

- **Modifying the transfer function** or its approximation to fit the experimental data. This involves computation of the coefficients (parameters) for the selected transfer function equation. After the parameters are found, the transfer function becomes unique for that particular sensor. The function can be used for computing the input stimuli from any sensor response within the range. Every calibrated sensor will have its own set of the unique parameters. The sensor is not modified.

Calibration

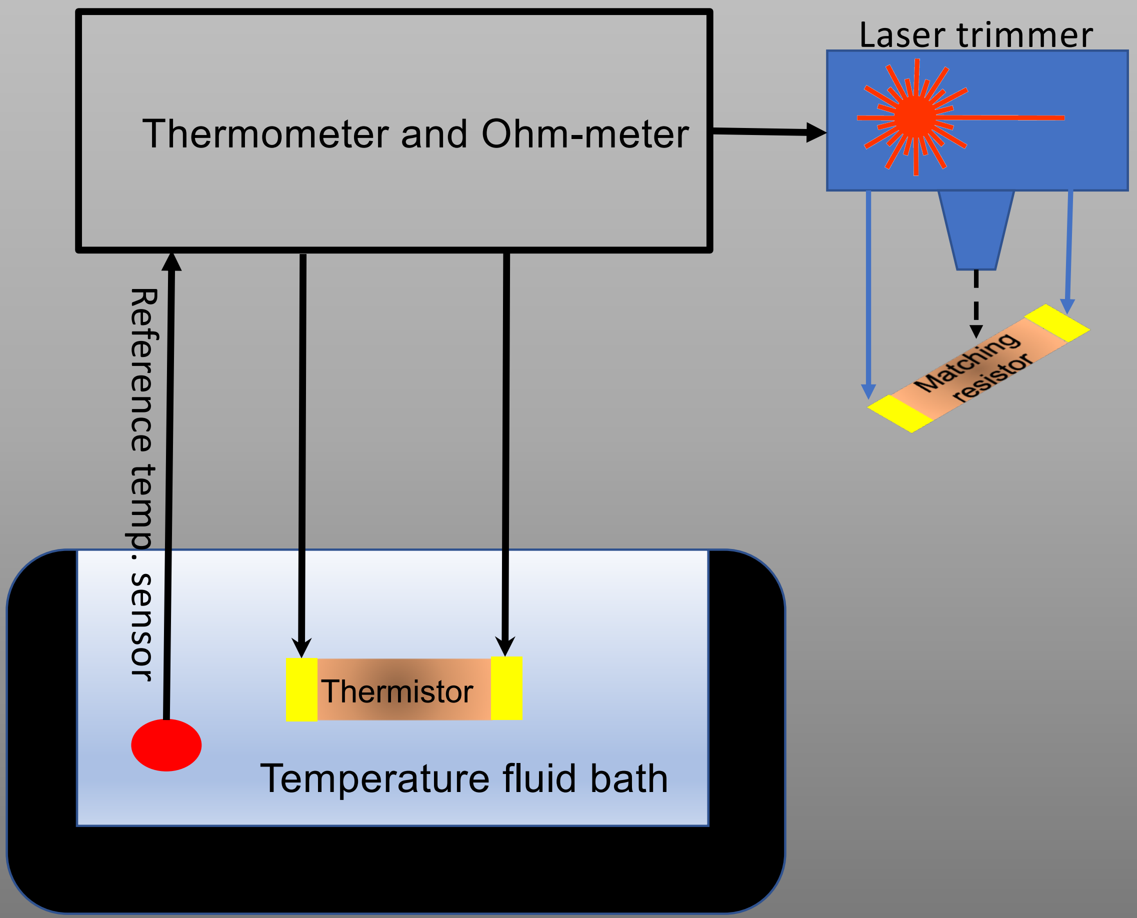
- **Adjustment of the data acquisition** system to trim (modify) its output by making the outputs signal to fit into a normalized or “ideal” transfer function. An example is a scaling and shifting the acquired data (modifying the system gain and offset). The sensor is not modified.
- Modification (trimming) the sensor’s properties to fit the predetermined transfer function, thus the sensor itself is modified.
- Creating the sensor-specific reference device with the matching properties at particular calibrating points. This unique reference is used by the data acquisition system to compensate for the sensor’s inaccuracy. The sensor is not modified.

Calibration example: option 3



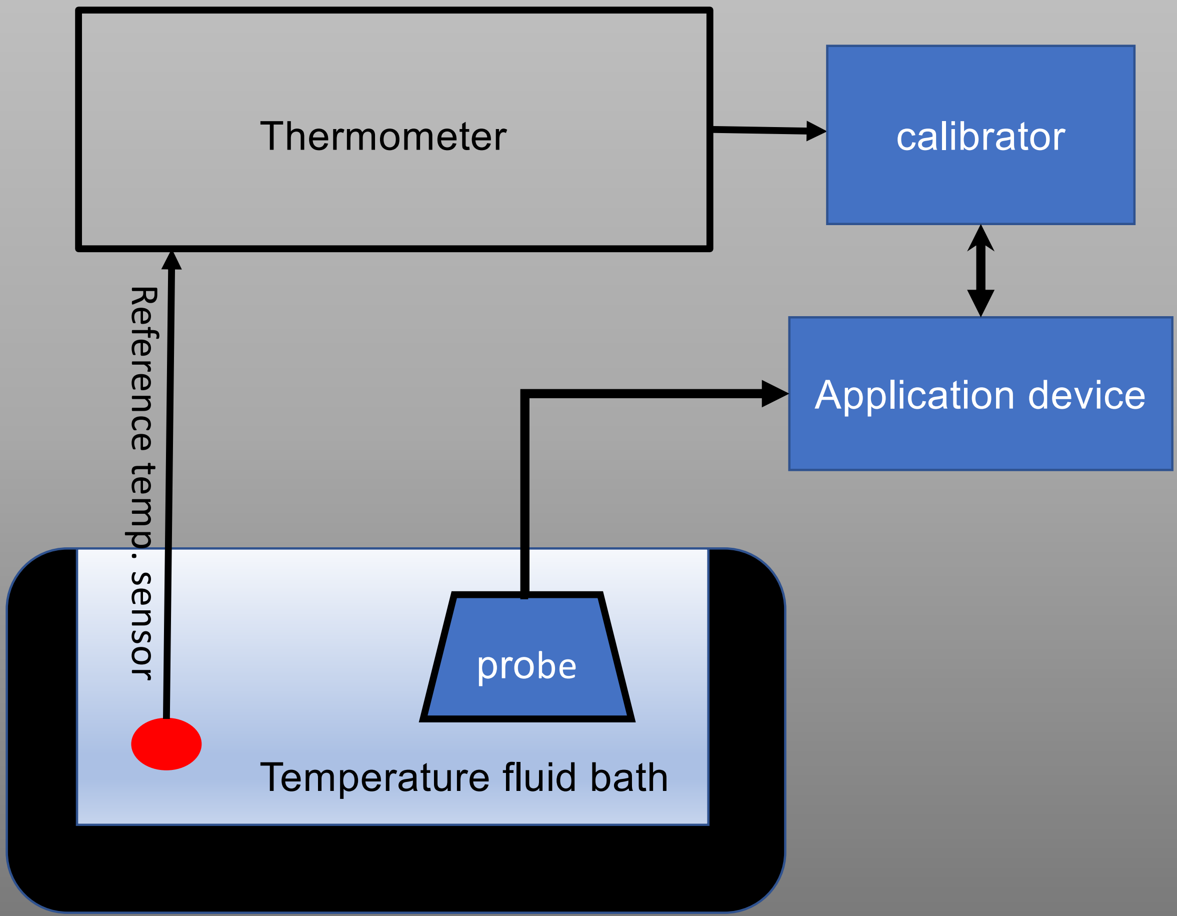
The figure shows a thermistor that is immersed into a stirred liquid bath with a precisely controlled and monitored temperature. The liquid temperature is continuously measured by a precision reference thermometer. To prevent shorting the thermistor terminals, the liquid should be electrically nonconductive, such as mineral oil or Fluorinert^{TM}. The resistance of the thermistor is measured by a precision Ohmmeter. A miniature grinder mechanically removes some material from the thermistor body to modify its dimensions (a). Reduction in dimensions leads to increase in the thermistor electrical resistance at the selected bath temperature. When the thermistor's resistance matches a predetermined value of the "ideal" resistance, the grinding stops and the calibration is finished. Now the thermistor response is close to the "ideal" transfer function, at least at that temperature. Naturally, a single-point calibration assumes that the transfer function can be fully characterized by that point.

Calibration example: option 2



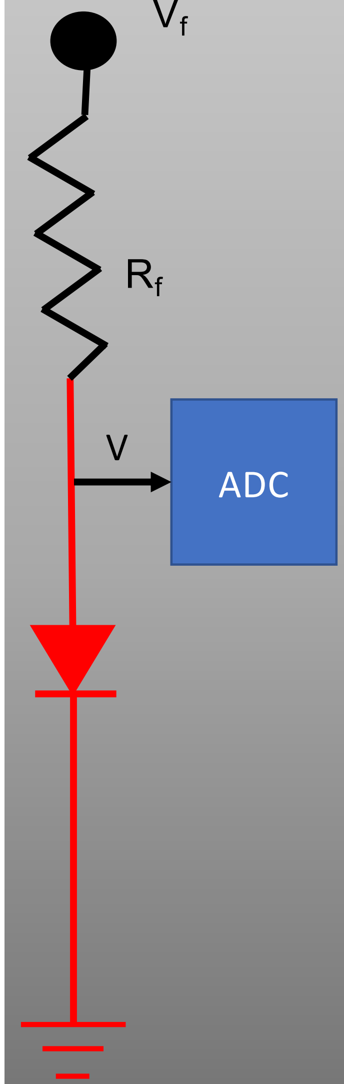
The measurement provides a number that is used for selecting a conventional (temperature stable) matching resistor as a unique reference. That resistor is for use in the interface scaling circuit. The precise value of such a reference resistor is achieved either by a laser trimming or selection from a stock. That individually matched pair thermistor–resistor is used in the measurement circuit, for example, in Wheatstone bridge. Since it is a matching pair, the response of the bridge will scale to correspond to an ideal transfer function of a thermistor.

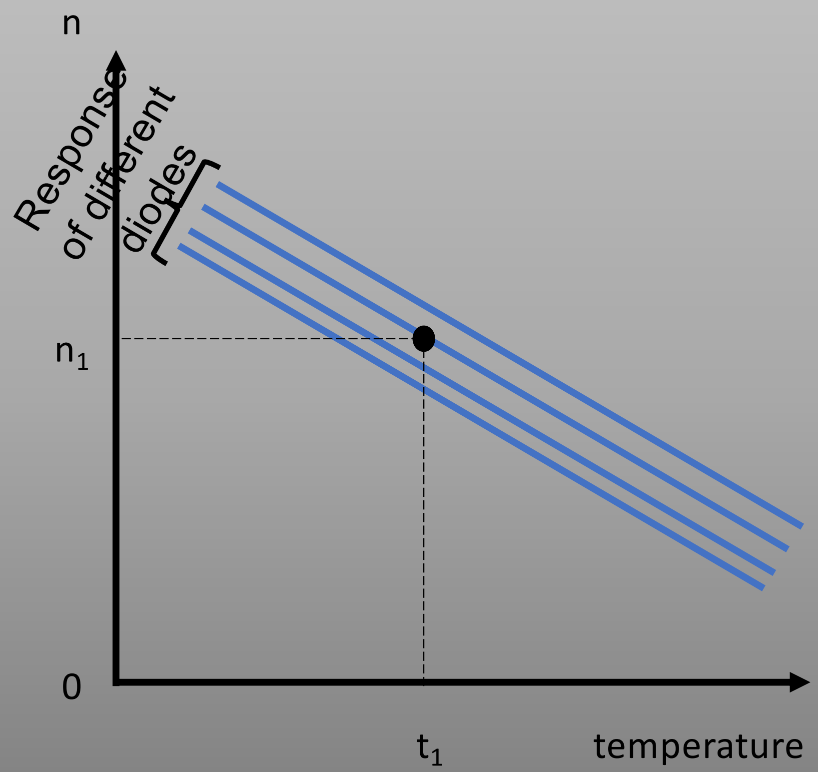
Calibration example: option 1



In the above examples, methods (3) and (2) are useful for calibration at one temperature point only, assuming that other parameters of the transfer function do not need calibration. If such is not the case, several calibrating points at different temperatures and resistances should be generated. Here, the liquid bath is sequentially set at two, three, or four different temperatures and the thermistor under calibration produces the corresponding responses, that are used by the calibrating device to generate the appropriate parameters for the inverse transfer function that will be stored in the application device (e.g., a thermometer).

Computation of parameters

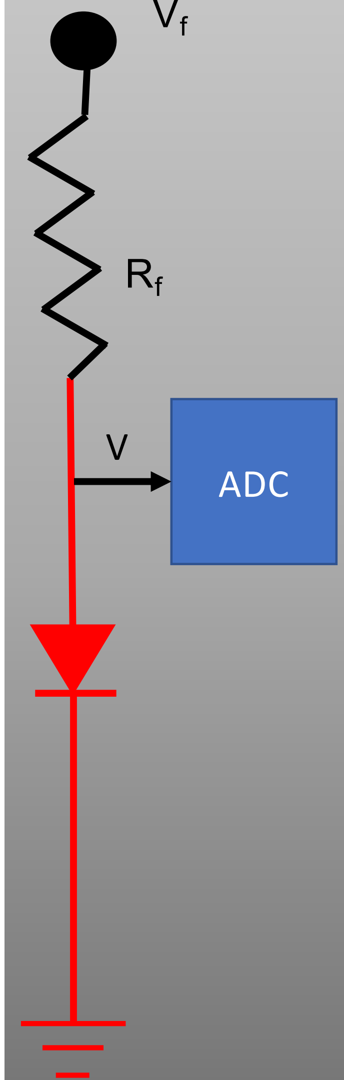


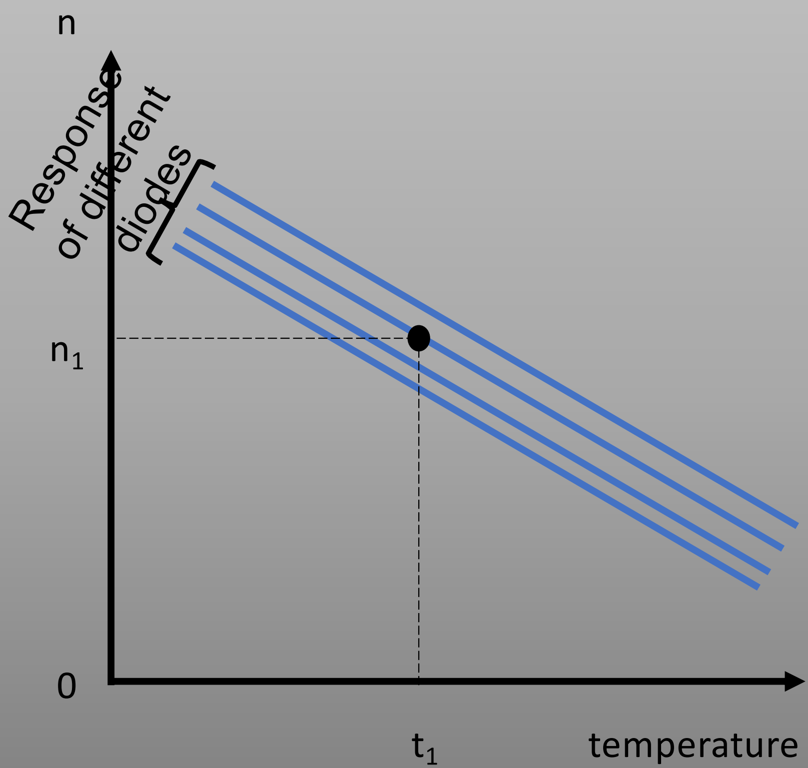


$$n(t) = n_1 + B(t - t_1)$$

If a transfer function is linear, then calibration should determine constants A and B. If it is exponential, the constants A and k should be determined, and so on.
To calculate parameters (constants) of a linear transfer function **one needs two data points defined by two calibrating input–output pairs**. Consider a simple linear transfer function, since two points are required to define a straight line, a two-point calibration shall be performed. For example, if one uses a forward- biased semiconductor p–n junction as a temperature sensor, its transfer function is linear (see figure) with temperature t being the input stimulus and the ADC count n from the interface circuit is the output:

Computation of parameters





$$n(t) = n_1 + B(t - t_1)$$

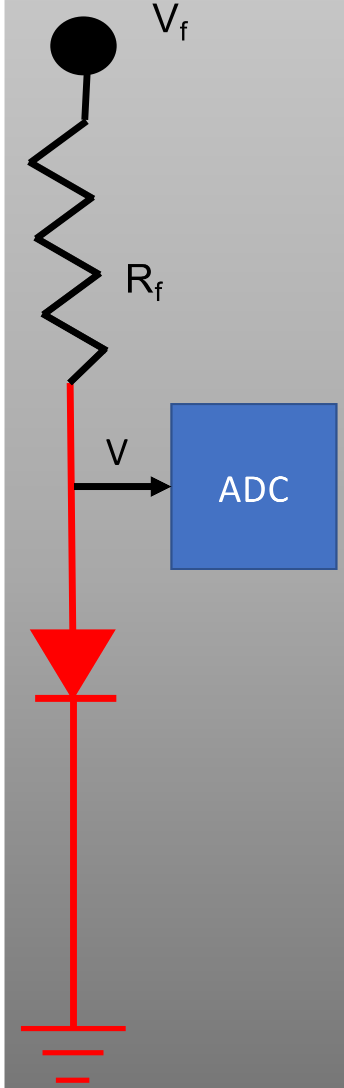
t_1 and n_1 are the coordinates of the first reference calibrating point. To fully define the line, the sensor shall be subjected to two calibrating temperatures (t_1 and t_2) for which two corresponding output counts (n_1 and n_2) will be registered. At the first calibrating temperature t_1, the output count is n_1.

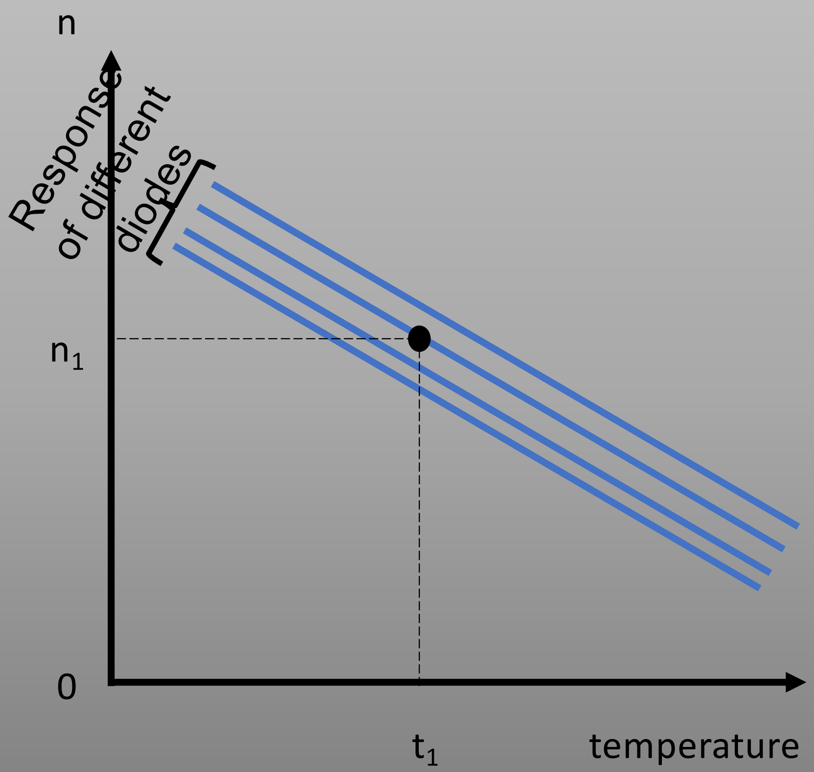
After subjecting the sensor to the second calibrating temperature t_2, we receive the digital counts for the second calibrating point. The count is

$$n_2 = n_1 + B(t_2 - t_1)$$

$$B = \frac{n_2 - n_1}{t_2 - t_1}$$

Computation of parameters



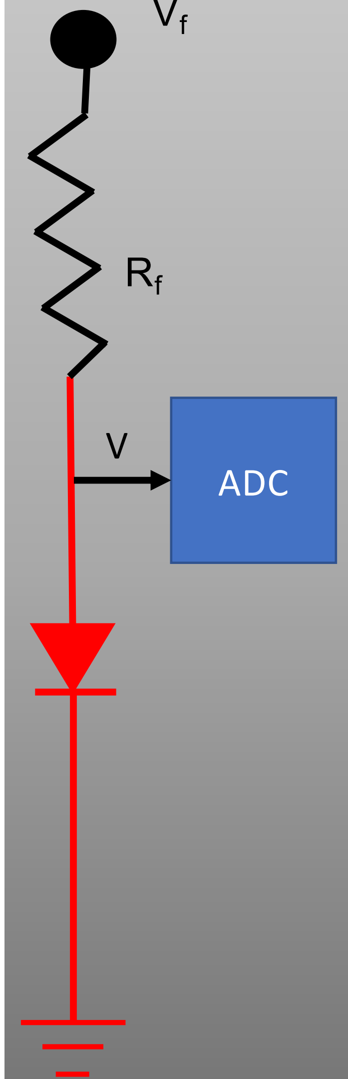


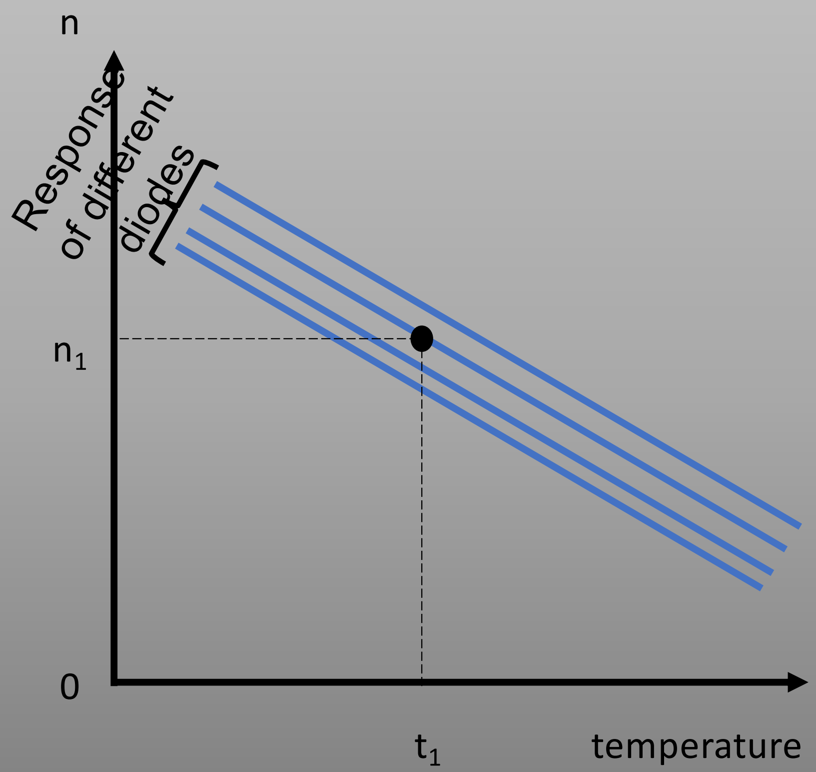
The sensitivity (slope) B is in count/degree. In the figure the slope B is negative since a p–n junction has a negative temperature coefficient (NTC). Note that the parameters found from calibration are unique for the particular sensor and must be stored in the measurement system to which that particular sensor is connected. For another similar sensor, these parameters will be different (perhaps except t_1, if all sensors are calibrated at exactly the same temperature). After calibration is done, any temperature within the operating range can be computed from the ADC output count n by use of the inverse transfer function

$$n(t) = n_1 + B(t - t_1)$$

$$t = t_1 + \frac{n - n_1}{B}$$

Computation of parameters





$$n(t) = n_1 + B(t - t_1)$$

In some fortunate cases, parameter *B* may be already known with a sufficient accuracy so that no computation of *B* is needed. In a p–n junction, the slope *B* is usually very consistent for a given lot and type of the semiconductor wafer and thus can be considered as a known parameter for all diodes in the production lot. However, all diodes may have different offsets, so a single-point calibration is still needed to find out n_1 for each individual sensor at the calibrating temperature t_1.

Computation of parameters

For nonlinear transfer functions, calibration at one data point may be sufficient only in some rare cases when other parameters are already known, but often two and more input–output calibrating pairs would be required. When a second or a third degree polynomial transfer functions are employed, respectively three and four calibrating pairs are required.

$$E = b_3 s^3 + b_2 s^2 + b_1 s + b_0$$

$$\begin{cases} E = b_3 {s_1}^3 + b_2 {s_1}^2 + b_1 s_1 + b_0 \\ E = b_3 {s_2}^3 + b_2 {s_2}^2 + b_1 s_2 + b_0 \\ E = b_3 {s_3}^3 + b_2 {s_3}^2 + b_1 s_3 + b_0 \\ E = b_3 {s_4}^3 + b_2 {s_4}^2 + b_1 s_4 + b_0 \end{cases}$$

Computation of parameters

$$\Delta = \left(\frac{s_1^2-s_2^2}{s_1-s_2} - \frac{s_1^2-s_4^2}{s_1-s_4}\right)\left(\frac{s_1^3-s_2^3}{s_1-s_2} - \frac{s_1^3-s_3^3}{s_1-s_3}\right) - \left(\frac{s_1^2-s_2^2}{s_1-s_2} - \frac{s_1^2-s_3^2}{s_1-s_3}\right)\left(\frac{s_1^3-s_2^3}{s_1-s_2} - \frac{s_1^3-s_4^3}{s_1-s_4}\right)$$

$$\Delta_a = \left(\frac{s_1^2-s_2^2}{s_1-s_2} - \frac{s_1^2-s_4^2}{s_1-s_4}\right)\left(\frac{E_1-E_2}{s_1-s_2} - \frac{E_1-E_3}{s_1-s_3}\right) - \left(\frac{s_1^2-s_2^2}{s_1-s_2} - \frac{s_1^2-s_3^2}{s_1-s_3}\right)\left(\frac{E_1-E_2}{s_1-s_2} - \frac{E_1-E_4}{s_1-s_4}\right)$$

$$\Delta_b = \left(\frac{s_1^3-s_2^3}{s_1-s_2} - \frac{s_1^3-s_3^3}{s_1-s_3}\right)\left(\frac{E_1-E_2}{s_1-s_2} - \frac{E_1-E_4}{s_1-s_4}\right) - \left(\frac{s_1^3-s_2^3}{s_1-s_2} - \frac{s_1^3-s_4^3}{s_1-s_4}\right)\left(\frac{E_1-E_2}{s_1-s_2} - \frac{E_1-E_3}{s_1-s_3}\right)$$

$$b_3 = \frac{\Delta_a}{\Delta} \qquad b_1 = \frac{1}{s_1 - s_4}\left[E_1 - E_4 - b_3(s_1^3 - s_4^3) - b_2(s_1^2 - s_4^2)\right]$$

$$b_2 = \frac{\Delta_b}{\Delta} \qquad b_0 = E_1 - b_3 s_1^3 - b_2 s_1^2 - b_1 s_1$$

If the determinant Δ is small, some considerable inaccuracy will result. Thus, the calibrating points should be spaced within the operating range as far as possible from one another.

Computation of parameters

A general goal of sensing is to determine the value of the input stimulus ***s*** from the measured output signal ***E***. This can be done by two methods.

1. From an inverted transfer function ***s = F(E)**, that may be either an analytical or approximation function,*
2. From a direct transfer function ***E*** = ***f(s)*** by use of an iterative computation.

Use of analytical equation

This is a straight approach when an analytical equation for the transfer equation is known. Simply measure the output signal ***E***, plug it into the formula, and compute the sought input stimulus s. For example, to compute a displacement from resistance of a potentiometric sensor, use

$$s = s_0 + \frac{E - E_0}{B}$$

For other functional models, use respective equations:

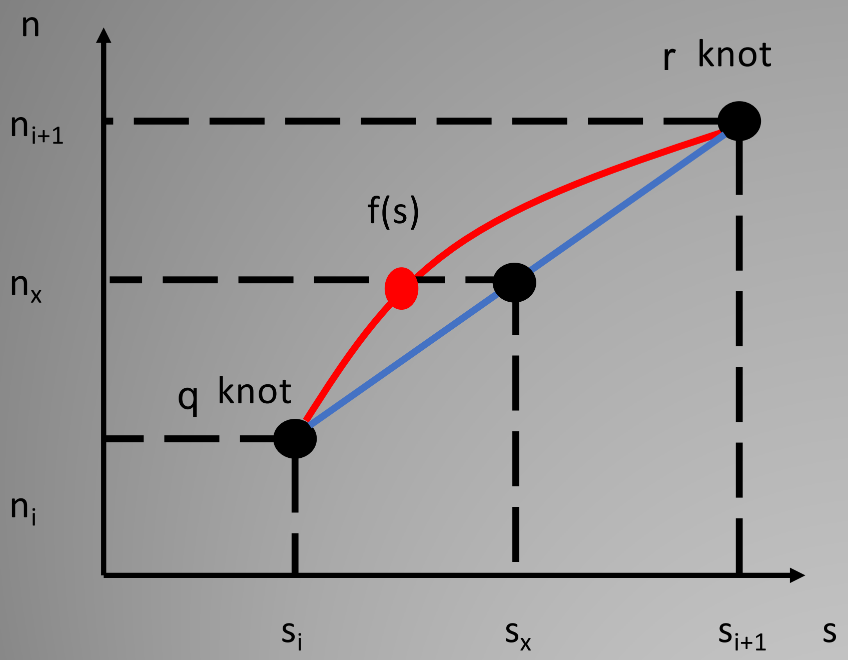
$$s = \frac{1}{k}\ln(E)$$

$$s = e^{\frac{E-A}{B}}$$

$$s = \sqrt[k]{\frac{E - A}{B}}$$

Use of linear piecewise approximation

For computing stimulus s, the very first step is to find out where it is located, in other words, between which knots lays the output signal ***E*** ? The next step is to use the method of linear interpolation for computing the input stimulus **s**.



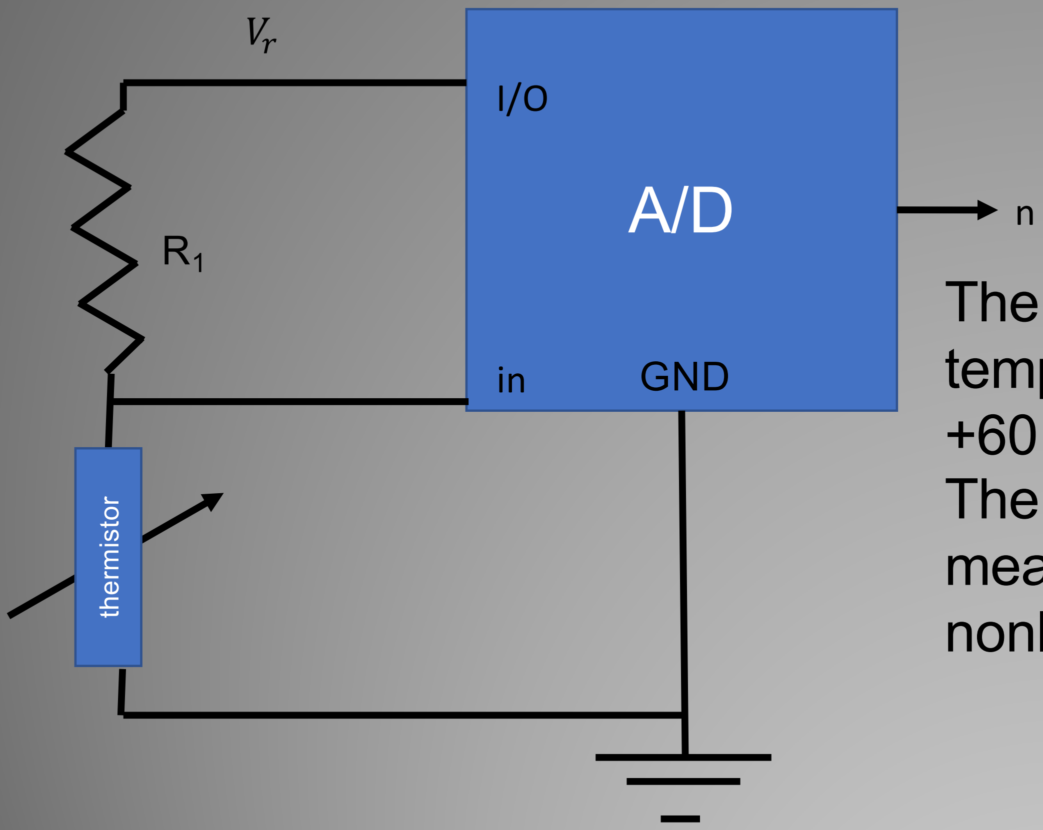
First, determine where the output is located, that is, in between which knots? For example, we found that the output is somewhere in between the knots q and r as illustrated in figure.

$$\frac{n_x - n_i}{n_{i+1} - n_i} = \frac{s_x - s_i}{s_{i+1} - s_i}$$

$$s_x = s_i + \frac{n_x - n_i}{n_{i+1} - n_i}(s_{i+1} - s_i)$$

This equation is easy to program and compute by an inexpensive microprocessor, which keeps in its memory a look-up table containing the knot coordinates

Use of linear piecewise approximation



ADC – 12 bit

Analog → Digital

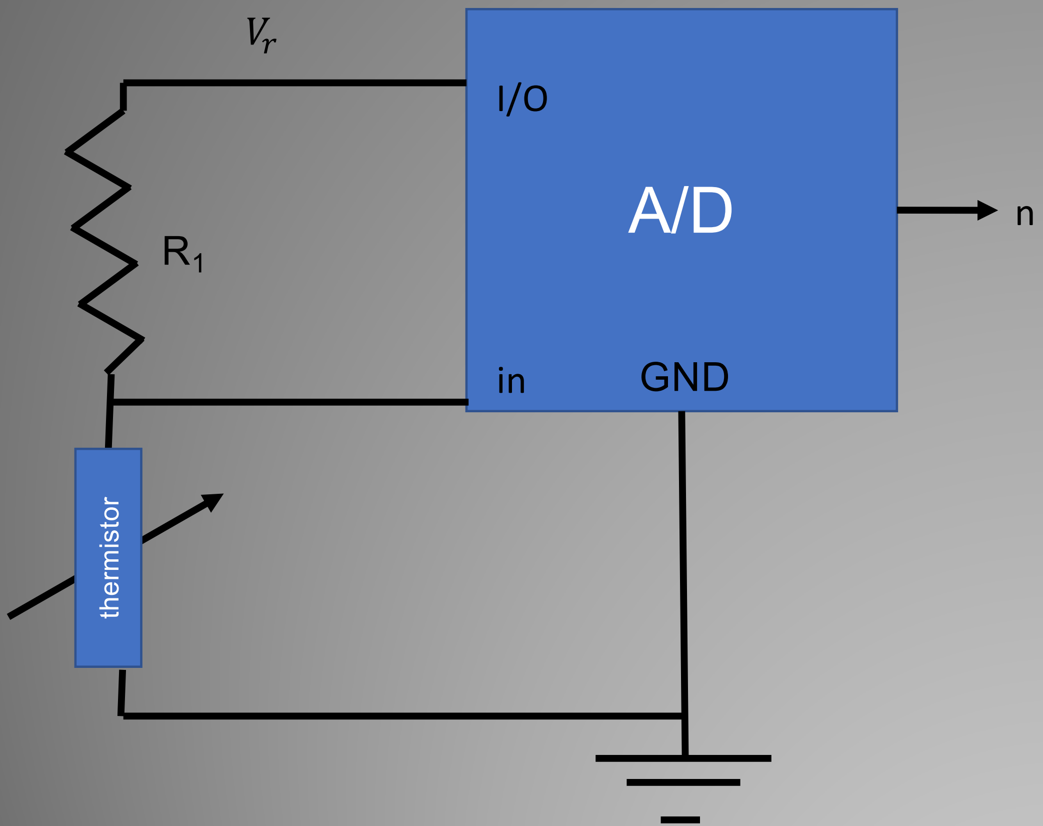
N=4095 counts ⟹ $V=V_r$

The thermistor is used to measure temperature in the total input span from 0 to +60 °C.

The output count of the thermistor measurements circuit can be modelled by a nonlinear function of temperature:

$$n_x = N_0 \frac{R_r e^{\beta\left(T_x^{-1} - T_r^{-1}\right)}}{R_1 + R_r e^{\beta\left(T_x^{-1} - T_r^{-1}\right)}}$$

Use of linear piecewise approximation



$$n_x = N_0 \frac{R_r e^{\beta\left(T_x^{-1} - T_r^{-1}\right)}}{R_1 + R_r e^{\beta\left(T_x^{-1} - T_r^{-1}\right)}}$$

where T_x is the unknown measured temperature, T_r is the reference temperature, R_r is resistance of the thermistor at reference temperature T_r, and β is the characteristic temperature. All temperatures and β are in degrees kelvin.

$$T_x = \left[\frac{1}{T_r} + \frac{1}{\beta} ln\left(\frac{n_x}{N_0 - n_x}\frac{R_1}{R_r}\right)\right]^{-1}$$

The above equation contain two unknown parameters: R_r and β. Thus, before we proceed further, the entire circuit, including the ADC, shall be calibrated at temperature T_r and also at some other temperature T_C. In the circuit, we use a pull-up resistor R_1 =10.0 kΩ. For calibration, we select two calibrating temperatures in the operating range as T_r = 293.15 K and T_c = 313.15 K, which correspond to 20 °C and 40 °C, respectively. In the circuit, we use a pull-up resistor R_1 =10.0 kΩ. For calibration, we select two calibrating temperatures in the operating range as T_r = 293.15 K and T_c = 313.15 K, which correspond to 20 °C and 40 °C, respectively.

During calibration, the thermistor sequentially is immersed into a fluid bath at these two temperatures and the ADC output counts are registered respectively as:

$n_r = 1863$ at T_r = 293.15 K

$n_c = 1078$ at T_r = 293.15 K

R_r=8.35 Ω

β = 3895 K

This complete the calibration.

Now, since all parameters are fully characterized, we can use the formula:

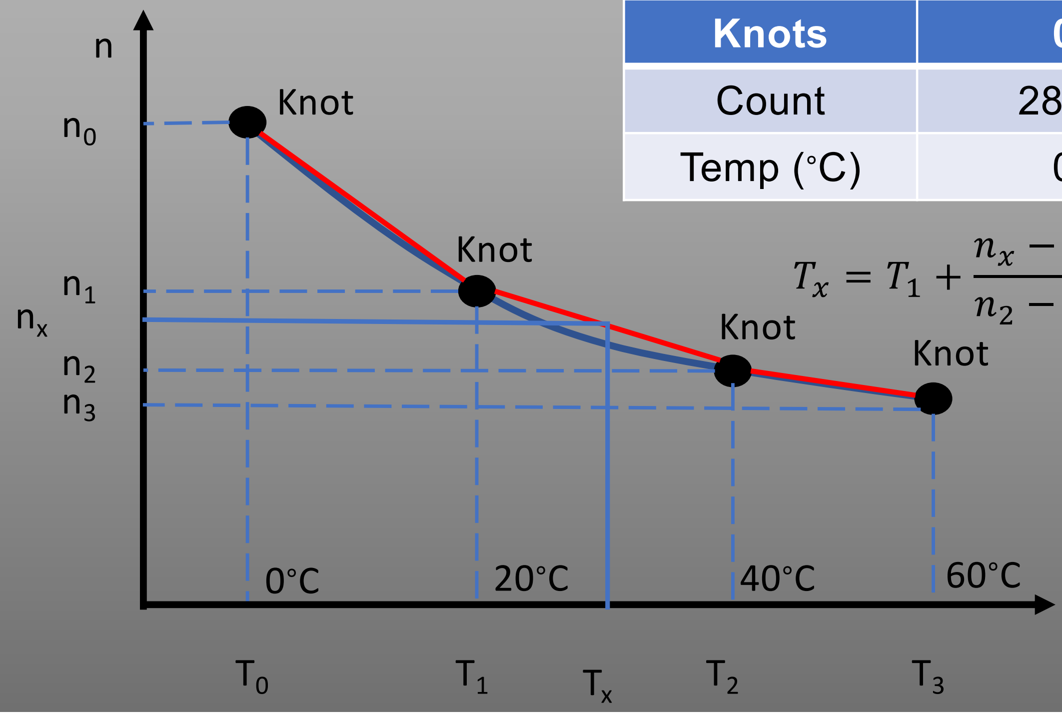
$$T_x = \left[\frac{1}{T_r} + \frac{1}{\beta} ln\left(\frac{n_x}{N_0 - n_x}\frac{R_1}{R_r}\right)\right]^{-1}$$

for computing temperature from any ADC count in the operating range. We assume this is the most accurate way of computing true temperature.

We assume this is the most accurate way of computing true temperature. We use now the linear piecewise approximation. We break up the transfer function just

$$n_x = N_0 \frac{R_r e^{\beta\left(T_x^{-1} - T_r^{-1}\right)}}{R_1 + R_r e^{\beta\left(T_x^{-1} - T_r^{-1}\right)}}$$

into three sections with two end knots at 0 and 60°C (the span limits) and two equally spaced central knots at 20 and 40° C.



Knots	0	1	2	3
Count	2819	1863	1078	593
Temp (°C)	0	20	40	60

$$T_x = T_1 + \frac{n_x - n_1}{n_2 - n_1}(T_2 - T_1) = 20 + \frac{1505 - 1863}{1078 - 1863}(40 - 20) = 29.12°C$$

$$n_x = N_0 \frac{R_r e^{\beta\left(T_x^{-1} - T_r^{-1}\right)}}{R_1 + R_r e^{\beta\left(T_x^{-1} - T_r^{-1}\right)}} = 28.22°C$$

$$\Delta = 0.9°C$$

Iterative computation of stimulus (Newton method)

If the inverse transfer function is not known, the iterative method allows using a direct transfer function to compute the input stimulus. A very powerful method of iterations is the Newton or secant method. It is based on **first guessing the initial reasonable value of stimulus** $s = s_0$ and then applying the Newton algorithm to compute a series of new values of s converging to the sought stimulus value. Thus, **the algorithm involves several steps of computation**, where each new step brings us closer and closer to the sought stimulus value. When a difference between two consecutively computed values of s becomes sufficiently small (less than an acceptable error), the algorithm stops and the last computed value of s is considered a solution of the original equation and thus the value of the unknown stimulus is found. Newton's method converges remarkably quickly, especially if the initial guess is reasonably close to the actual value of s.

The output signal is represented through the sensor's transfer function is f(s) as **E** = f(**s**). It can be rewritten as **E** - f(**s**) = 0. The Newton method prescribes computing the following sequence of the stimuli values for the measured output value E:

$$s_{i+1} = s_i - \frac{f(s_i) - E}{f'(s_i)}$$

This sequence after just several steps converges to the sought input s. Here, s_{i+1} is the computed stimulus value at the iteration i + 1, wherein s_i is the computed value at a prior iteration i and $f'(s_i)$ is the first derivative of the transfer function at input s_i. The iteration number is i = 0, 1, 2, 3, ... Note that the same measured value E is used in all iterations.

Iterative computation of stimulus (Newton method)

Start by guessing stimulus s_0, then

$$s_{i+1} = s_i - \frac{f(s_i) - E}{f'(s_i)}$$

Allows to calculate the next approximation to the true stimulus s. Then, do it again by using the result from the prior approximation of s. In other words, computation of the subsequent s_i is performed several times (iterations) until the incremental change in s_i becomes sufficiently small, preferably in the range of the sensor resolution. To illustrate use of the Newton method let us assume that our direct transfer function is a third degree polynomial:

$$f(s)=as^3+bs^2+c^s+d$$

having coefficients a = 1.5, b = 45, c= 25, d= 1.

$$s_{i+1} = s_i - \frac{as_i^3+bs_i^2+cs_i+d-E}{3as_i^2+2bs_i+c} = \frac{2as_i^3+bs_i^2-d+E}{3as_i^2+2bs_i+c}$$

Iterative computation of stimulus (Newton method)

$$s_{i+1} = s_i - \frac{as_i^3+bs_i^2+cs_i+d-E}{3as_i^2+2bs_i+c} = \frac{2as_i^3+bs_i^2-d+E}{3as_i^2+2bs_i+c}$$

E=22.0 and s_0=2

$$s_1 = \frac{2\cdot1.5\cdot2^3+5\cdot2^2-1+22}{3\cdot1.5\cdot2^2+2\cdot5\cdot2+25} = 1.032$$

$$s_2 = \frac{2\cdot1.5\cdot1.032^3+5\cdot1.032^2-1+22}{3\cdot1.5\cdot1.032^2+2\cdot5\cdot1.032+25} = 0.738$$

$$s_3 = \frac{2\cdot1.5\cdot0.738^3+5\cdot0.738^2-1+22}{3\cdot1.5\cdot0.738^2+2\cdot5\cdot0.738+25} = 0.716$$

$$s_4 = \frac{2\cdot1.5\cdot0.716^3+5\cdot0.716^2-1+22}{3\cdot1.5\cdot0.716^2+2\cdot5\cdot0.716+25} = 0.716$$

It should be noted that the Newton method results in large errors when the sensor's sensitivity becomes low. **In other words, the method will fail where the transfer function flattens** (1st derivative approaches zero). In such cases, the so-called Modified Newton Method may be employed.

$$f(0.716) = 22.014$$

0.06%